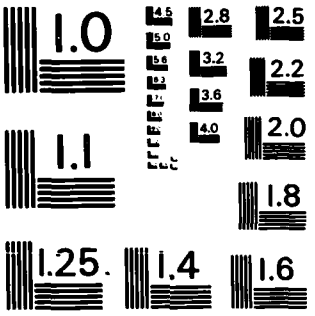
AD-A148 461 AUTOMATIC SPEECH RECOGNITION IN SEVERE ENVIRONMENTS(U) 1/1
NATIONAL RESEARCH COUNCIL WASHINGTON DC COMMITTEE ON
COMPUTERIZED SPEECH RECOGNITION TECHNOLOGIES 1984
UNCLASSIFIED N00014-83-G-0072 F/G 5/8 NL



AUTOMATIC SPEECH RECOGNITION IN SEVERE ENVIRONMENTS

A Report Prepared by the

Committee on Computerized Speech Recognition Technologies
Commission on Engineering and Technical Systems
National Research Council

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A148 461
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Automatic Speech Recognition in Severe Environments	5. TYPE OF REPORT & PERIOD COVERED Final Report: 15 Apr 83 - 15 Aug 84
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Committee on Computerized Speech Recognition Technologies	8. CONTRACT OR GRANT NUMBER(s) N00014-83-G-0072
9. PERFORMING ORGANIZATION NAME AND ADDRESS National Research Council 2101 Constitution Avenue, N.W. Washington, D.C. 20418	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N 42; RR04209; RR0420901; NR 196-181
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 442) 800 N. Quincy Street Arlington, VA 22217 - 5000	12. REPORT DATE 1984
	13. NUMBER OF PAGES 14 pp.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same	15. SECURITY CLASS. (of this report) U
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

For public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

computerized speech recognition technologies; automatic speech recognition; voice input/output; speech synthesis; voice interactive

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The report examines the status and outlook for voice interaction in moderate and severe environments; the primary focus is on the technology of speech recognition and human factor issues associated with its application. Issues in the technology voice input/output are reviewed, and basic concepts in speech technology are investigated. A current status report on computerized speech technology is given and current applications are discussed. The committee indicates directions for future work and then ends with conclusions and recommendations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

INTRODUCTION

For more than a decade the United States government, private corporations, and universities have been engaged in research on human-machine interaction by voice. The Department of Defense (DOD), in particular, has recognized the potential for improving the safety and effectiveness of its forces by making electronic and electromechanical devices directly responsive to the human voice and able to respond by voice. The benefits of this capability would be especially noteworthy in situations where the individual is engaged in such hands/eyes-busy tasks as flying an airplane or operating a tank.

In order to provide a forum for the exchange of ideas and information, and, additionally, to coordinate the activities of several federal agencies and research groups dealing with speech input/output systems, the Department of Defense established a Voice SubTechnical Advisory Group. This group functions within the Defense Department's Human Factors Technical Advisory Group (TAG). Thus, members of the Voice SubTAG are some of the government researchers and program managers who must respond to the increasing pressures to demonstrate and deliver computer-based, voice-interactive systems capable of reliable operation in severe and moderate military environments.

In view of these responsibilities, the Voice SubTAG asked the National Research Council to conduct an independent evaluation of the status and outlook for voice-interaction in moderate and severe environments. Recognizing also that the technology of speech synthesis is further advanced, the request to the National Research Council centered specifically upon the technology of speech recognition and upon the human factors issues associated with the application of speech recognition to government and military tasks. Speech synthesis is addressed, in this report, only in the context of the human factors issues associated with the design of voice-interactive systems.

The study plan contained six objectives:

1. Identify major applications of automatic speech recognition that have existing or potential utility in both stressful and moderate environments characterized by both moderate and severe levels of psychological and physiological stress.

2. Review and summarize the properties of human speech under stressful conditions and in the presence of different types of background noise characteristic of the envisaged settings.
3. Review and summarize the state of the art and expected progress in speech recognition algorithms and systems, including procedures for enhancing information flow and rejecting noise.
4. Recommend the nature of performance guidelines and standards needed for further development and application.
5. Outline performance tradeoffs and assessment procedures, including consideration of the "systems cost" of the additional information channel offered by computerized speech recognition and synthesis.
6. Identify required research in fundamental areas of speech technology and its applications.

Responding to this request, the National Research Council's Commission on Engineering and Technical Systems assembled the Committee on Computerized Speech Recognition Technologies. The committee was comprised of eleven specialists from fields pertinent to speech recognition. These fields included speech communication, computer science, electrical engineering, human factors, acoustics, psycholinguistics, avionic systems, natural language processing, and artificial intelligence. The committee was drawn from industry and academia. The committee's task was to make projections and recommendations for future development of speech recognition technology, particularly for the military sector.

The committee met five times, for approximately two days each time, from July 1983 to March 1984. The initial meetings were devoted to information gathering. The later meetings were devoted to analysis of the information, specifically as it related to government applications, and to formulating this written report. Additionally, the results of the committee's study were presented to the committee's sponsors in a briefing on August 2, 1984.

In the course of its study, the committee:

- was briefed by members of the Voice SubTAG on the content of their programs and their expectations regarding the committee's role;
- heard presentations by invited consultants and experts on the state of the art of laboratory research and commercial systems;
- discussed the goals, procedures, and accomplishments of DOD's Advanced Research Projects Agency's Speech Understanding Research Project in the 1970s; and
- examined possible applications of speech recognition systems in the military sector; for example, in fighter aircraft and shipboard command/control centers.

These investigations involved briefings by equipment vendors and systems designers, interviews with test pilots, and field trips to development laboratories, test facilities, and operations environments.

This report should be of use to DOD and civilian policy makers who determine the capabilities that will be required in advanced combat systems. For these readers, the report gives the committee's best judgment regarding the probability of meeting various requirements in the future. Laboratory researchers and research managers may also look to the report for guidance in planning future activities.

The committee's ability to meet its obligations was heavily dependent upon assistance provided by a number of interested and accommodating parties. We are deeply appreciative of their kindness and efforts.

In particular, we thank Colonel Harry Heimple, director of the Advanced Fighter Technology Integrator (AFTI) F/16 Program, and all other members of his team, including test pilots, equipment vendors, General Dynamics staff, and National Aeronautics and Space Administration (NASA) staff for hosting the site visit to Edwards Air Force Base, California. We thank Captain James R. Williams, commanding officer of the Naval Surface Weapons Center and the members of his staff for briefing the committee and providing a tour of Systems Control Laboratory at Dahlgren, Virginia. We thank Leon Lerman of Lockheed Missiles and Space Company, Sunnyvale, California, for providing his first-person account of lessons learned in applying speech recognition systems to everyday industrial operations.

Additionally, we thank the commanding officer of the cruiser U.S.S. Long Beach, Captain Frederick Triggs, USN, his executive officer, Captain Clyde J. Vanarsdall, USN, and Master Chief Jess Mahon and other crew members for a memorable and productive tour of the ship's combat information center.

The committee values and appreciates the interest and counsel that it has received from its two liaison members from the Commission on Engineering and Technical Systems--C. Kumar N. Patel and Erich Bloch.

Finally, the committee wishes to recognize the efforts of Dennis F. Miller, study director, in developing and organizing this study and, for his guidance throughout this performance period. Further, the committee wishes to express its appreciation to Howard Clark, staff officer, for his assistance. Also, the committee wants to thank Helen Johnson, Patricia Wood, June Richardson, and Julia Torrence for their cheerful help and generous support.

SUMMARY OF REPORT

COMMITTEE ON COMPUTERIZED SPEECH RECOGNITION TECHNOLOGIES (1984)
AUTOMATIC SPEECH RECOGNITION IN SEVERE ENVIRONMENTS
WASHINGTON, D.C.: NATIONAL ACADEMY PRESS

HUMAN-MACHINE INTERACTION BY VOICE

Speech is a natural and convenient means for human communication. Information exchange between humans and complex machines could be facilitated if machines could respond appropriately to spoken commands through action, information processing, or machine-generated voice. In particular, interaction by voice could alleviate the information load on the human in many "hands-busy/eyes-busy" situations.

Heavy information loads place severe demands on personnel such as pilots of single-seat, high-performance aircraft, tank crews in the field, and combat information center staff. Voice control of navigational displays, information files, and weapons systems could relieve the information loads on visual and manual channels.

The techniques of automatic speech recognition allow the machine to respond to spoken commands. The techniques of speech synthesis permit the machine to generate spoken responses. Over the past decade notable advances have been made in the development of automatic speech recognition and speech synthesis devices. These advances have been fueled by explosive progress in integrated circuit technology. Rudimentary systems are now being commercialized and more sophisticated techniques are under laboratory study.

But present speech recognition systems are rigidly limited in capability and require significant complexity for reliable performance. And, most speech systems have been used by forgiving users in benign environments. In contrast, military applications often involve harsh environmental conditions and demanding tasks where humans may be exposed to high ambient acoustic noise, encumbered by equipment (such as an oxygen or gas mask), and subjected to significant physiological and psychological stress during combat conditions.

Applications of speech synthesis are presently more extensive than those of recognition because message-generation techniques are somewhat better understood. Speech synthesis has been, and will

continue to be, a vital adjunct to recognition for voice-interactive systems, and its development must go apace. But the challenges in automatic recognition are greater, and this report focuses primarily on the issues of speech recognition.

ORIGIN OF THE PRESENT STUDY

Several military and government organizations requested that the National Research Council organize a study of the status and outlook for automatic speech recognition. The study was organized to consider the level of performance that present speech technology can support under hostile conditions, and what research and development might be undertaken to create a speech technology that could serve usefully in severe environments. The study aims to complement and expand the scope of previous efforts, such as the assessment conducted by Beek, Neuburg, and Hodge in 1977.

COMMITTEE ON COMPUTERIZED SPEECH RECOGNITION TECHNOLOGIES

In response to this request, the National Research Council convened a committee of eleven specialists to make projections and recommendations for future development of speech recognition technology, particularly for the military sector. The committee met five times, for approximately two days each time, over the interval July 1983 to March 1984. The initial meetings were devoted to information gathering--from presentations by the sponsors, current contractors, invited consultants and experts, and from several site visits to military installations. The later meetings were devoted to analysis of the information, specifically as it related to government applications, and to developing this written report. Additionally, in August 1984, the committee conducted an open briefing for its sponsors and other interested parties.

CONCLUSIONS

Based on its exposure to the issues and its familarity with the field, the committee concluded* that:

- The use of speech for communication between humans and machines has distinct potential for military and other government purposes.
- Current technology for automatic speech recognition is not sufficiently advanced to provide robust, reliable performance in hostile and high-stress environments.
- Current speech recognition technology is not sufficiently advanced to achieve high performance on continuous spoken input with large vocabularies and/or arbitrary talkers.
- Current technology is mature enough to support restricted applications in benign environments, with disciplined use under low-stress conditions. Success strongly depends upon the integration of speech recognition with improved automation techniques.
- No standardized techniques exist for evaluating and comparing the performance of speech recognizers.
- No established human-factors methodologies exist for analyzing and evaluating human-machine performance in integrated voice-interactive systems or for systematically quantifying the benefits of speech input as compared to related automation techniques.
- There is insufficient fundamental understanding of how human speech degrades under severe environmental and stress conditions and of how to design recognition algorithms for these conditions.
- Government-sponsored efforts are currently insufficient to sustain major advances in speech recognition technology.
- Laboratory studies of speech recognition algorithms will probably require sophisticated computational resources that are not widely available.
- Successful deployment of advanced speech recognition systems will be directly related to, and in part dependent upon, continued advances in integrated circuit technology and computer architecture.
- Speech synthesis is an important adjunct to automatic speech recognition for voice-interactive systems.
- No central focus exists in the U.S. government to manage research and development in speech recognition.

*Abstracted from detailed conclusions in Chapter 6.

RECOMMENDATIONS

The committee's conclusions lead to corollary recommendations. These recommendations* aim to achieve a speech recognition technology that can provide utility, accuracy, and reliability in severe as well as benign environments.

- A basic research program is needed to characterize speech and its variabilities, including the study of the acoustic properties of speech in various contexts and for different speakers.
- Because an automatic speech recognizer is limited by the information delivered to it, new methods for sound transduction (including microphone systems designed for severe environments) and for electronic signal enhancement should be sought and studied.
- Significant research efforts are required in the design of algorithms and systems for the recognition of continuous speech in complex application domains, for speaker-independent operation, and for robust performance under conditions of degraded input.
- Research is necessary to establish human-factors procedures for analyzing human-machine communication tasks, to quantify the benefits that speech input/output can contribute, and to develop systematic techniques for integrating speech functions into the systems design.
- Extensive hardware development and deployment, based on existing technology, is inappropriate. Exploratory hardware efforts, however, are vital for gaining practical knowledge about applications and for establishing the limitations of existing technology.
- Standardization should be established to quantify the performance of automatic speech recognizers and to permit comparisons among algorithm philosophies and environments. Common data bases and prescribed procedures for assessing performance should be made generally available.
- Sophisticated computational capabilities are required to support continued advances in speech recognition work. A program for advanced research in speech recognition should have appropriate interfaces with government-sponsored work on high-speed processors and strategic computing.
- A substantial, sustained, and coordinated program of research and development is required to realize the potential of speech recognition within the U.S. government. The program should be built around long-range goals, with the acquisition of fundamental knowledge as a central thrust. This objective is especially crucial to advancing continuous speech recognition and to achieving talker independence with large vocabularies. A focus of responsibility and accountability as well as a means for coordinating the program is necessary.

*Abstracted from detailed recommendations presented in Chapter 6.

AUTOMATIC SPEECH RECOGNITION WORK OUTSIDE THE UNITED STATES

At present, about two dozen companies are marketing automatic speech recognition devices in the United States. Their products are based on technology that largely originated in a few industrial laboratories and universities. The majority of this fundamental technology is available in the open technical literature, although some is protected by relatively narrow patents, and some is retained as trade secrets.

A similar fundamental understanding and capability exists in several other countries. In particular, significant expertise resides in Japan, Sweden, France, England, West Germany, and Canada. Of these, Japan has the most advanced effort, possibly equaling or surpassing that of the United States. Their motivation for automatic speech recognition (ASR) is double-edged. First, the syllabic structure of spoken Japanese is amenable to automatic speech recognition. Second, the complexities of the written language (Kanji and Kana characters) make keyboard design and operation burdensome, and make automatic "voice-to-text" conversion highly desirable.

Two significant forces have been at work in Japan to foster early ASR applications. One is the national telecommunications industry, led by Nippon Telephone and Telegraph (NTT) in partnership with its "family" companies (such as Nippon Electric Company, Fujitsu, Hitachi, Mitsubishi, Oki, and, more recently, Toshiba and Matsushita). The other is government stimulation by the Ministry of International Trade and Industry (MITI) of the "fifth generation" computer research, which has a strong component of voice input/output capability. As a result, the Japanese now have 20 to 40 speaker-independent, isolated-word systems working in their public telephone network. Each system, located in a telephone central office, can serve up to 128 calling lines. These small vocabulary systems serve the Japanese banking and financial industry. No comparable field deployment of ASR exists in other countries. Additionally, a number of simpler, lower-cost ASR systems are becoming available for personal computers, factory data entry, and device control. The ASR systems are typically combined with speech synthesis components based on technology that ranges from 32k bits/s Adaptive Differential Pulse Code Modulation (ADPCM) down to 2.4k bits/s Linear Predictive Coding (LPC). These technologies are well understood and in advanced states in Japan.

In France, ASR work is supported by the National Telephone Research Laboratory (CNET) and by the National Center for Scientific Research (CNRS) in their LIMSI laboratory. The results of this national research are typically available under license to industry. Some initial, though not widespread, commercialization has been made of the CNRS output. Applications under development include voice control of systems in the Mirage fighter aircraft and of power windows, locks, and directional signals as options in automobiles.

In Sweden, most recognition and synthesis products stem from research conducted in the Speech Transmission Laboratory of the Royal Technical University (KTH). Small venture businesses are selling text-synthesis and word recognition products. The developments also are being fostered by the Swedish government as communication aids for the handicapped.

Experimental products in England, West Germany, and Canada seem to be following very much the lines of the U.S. work, with British Telecom and Marconi, Siemens and Philips, and Bell Northern, respectively, included among the participants.

Speech recognition work has a long history in the Soviet Union, but the paucity of sophisticated computers and advanced electronics for civilian research seems to have retarded its development. At least on the basis of open publications, their progress seems several years behind the West.

In summary, while the United States is at the forefront in fundamental understanding and electronic implementation, it is not the undisputed leader in speech recognition and synthesis. Much of the developed world has comparable knowledge; Japan, at least, may have superior implementation. The competition is close enough, however, that modest additional investment in fundamental research could give the United States a significant edge.

BRIEFINGS

Technical briefings provided much of the information collected by the Committee on Computerized Speech Recognition Technologies. However, realizing there is no substitute for first-hand experience, members of the committee participated in a number of field trips. In this way, the committee was able to improve its understanding of the specialized environments and stressful situations in which computerized speech recognition systems must function. Additionally, members interacted directly with more equipment users and researchers than would otherwise have been possible.

These field trips are summarized below:

1. July 21, 1983 - A panel composed of committee members George Doddington, Ellen Roland, and John Ruth, plus staff officer Howard Clark, visited Edwards Air Force Base in California to observe flight tests of an operational speech recognition system in an AFTI F-16 aircraft. The panel received briefings on the program's goals and accomplishments, got a close-up view of the aircraft cockpit, and interviewed three test pilots.
2. September 15, 1983 - The committee visited the System Control Laboratory of the Naval Surface Weapons Center in Dahlgren, Virginia, and received briefings on the problems and potential applications of speech recognition systems in the surface Navy.
3. October 19, 1983 - Committee member Carol Simpson visited Lockheed Missiles and Space Company in Sunnyvale, California, to learn about their use of commercial voice recognition systems for data entry of part numbers, dash numbers, lot numbers, and serial numbers to provide traceability of parts--a service that is required of military suppliers. Lockheed has used voice technology data-entry for five years in its hybrid circuit assembly section and for three years in an electronic assembly area.

4. October 26, 1983 - The committee's staff officer, Howard Clark, toured the Army's AVRADA research facilities in Fort Monmouth, New Jersey. Current research activities there involve a performance evaluation of various commercial speech recognition systems operating in a simulated (noisy) helicopter environment. In another laboratory, Army researchers have installed a speech recognition system in a helicopter simulator and are examining its potential mission applications.
5. November 8, 1983 - The committee toured the Combat Information Center of the guided missile cruiser, U.S.S. Long Beach. Committee members were able to explore the working environment and procedural constraints within which a speech recognition system must function.
6. November 11, 1983 - Committee member Carol Simpson went to sea aboard the U.S.S. Long Beach to witness the operation of the Combat Information Center under more realistic conditions than were afforded during the committee's visit on November 8, 1983.

COMMITTEE ON COMPUTERIZED SPEECH RECOGNITION TECHNOLOGIES

James L. Flanagan (Chairman), Head, Acoustics Research Department, AT&T Bell Laboratories

N. Rex Dixon, Associate Editor, IBM Journal of Research and Development, IBM Corporation

George R. Doddington, Manager, Speech Systems Research Branch, Computer Science Laboratory, Texas Instruments

John I. Makhoul, Principal Scientist and Manager, Speech Signal Processing Department, Bolt Beranek & Newman Inc.

Michael E. McCauley, Vice President, Monterey Technologies, Inc.

Ellen F. Roland, Co-owner, Rolands and Associates Corporation

John C. Ruth, Vice President - Marketing and New Business Development, McDonnell Douglas Electronics Company

Carol A. Simpson, Co-owner, Psycho-Linguistic Research Associates

Beverly H. Williges, Senior Research Associate, Department of Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University

William A. Woods, Chief Scientist, Applied Expert Systems, Inc.

Victor W. Zue, Assistant Professor, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

---

COMMISSION ON ENGINEERING AND TECHNICAL SYSTEMS
LIAISON MEMBERS

Erich Bloch, IBM Vice President - Technical Personnel Development

C. Kumar N. Patel, Executive Director - Research Physics Division, AT&T Bell Laboratories

---

STAFF

Dennis F. Miller, Study Director
Howard E. Clark, Staff Officer
Helen D. Johnson, Staff Associate
June F. Richardson, Administrative Secretary

END

FILMED

1-85

DTIC